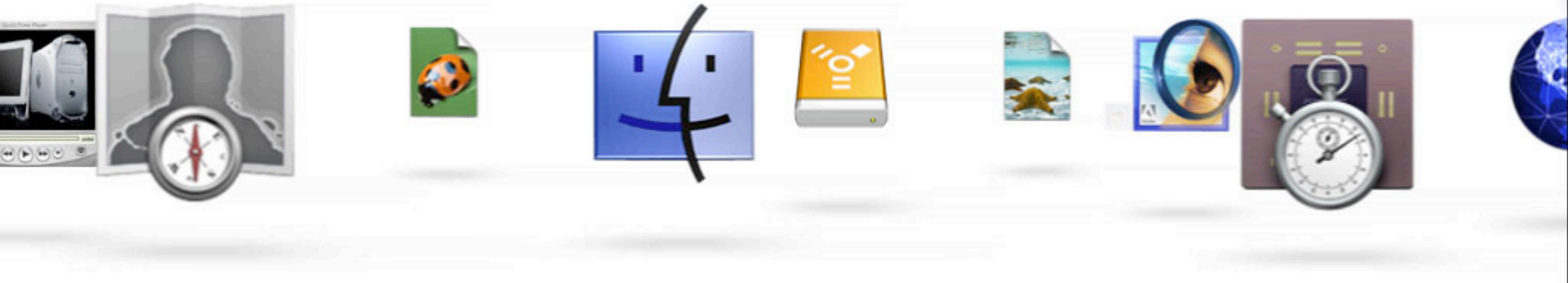




Speech Technologies in Mac OS X

Session 013





Speech Recognition and Synthesis In-Depth

John Geleynse
User Experience Evangelist

Agenda

- Why Adopt Speech?
- What's New In Jaguar?
- Speech Development Scenarios
 - What should I do first when adopting Speech?
 - Good
 - Better
 - Best





Speech Recognition and Synthesis In-Depth

Kim Silverman, PhD
Principal Research Scientist
Manager, Spoken Language Technologies



Demo

Shipping Features

What's New In Jaguar?

- Improvements in
 - Recognition
 - Synthesis
 - User Experience
 - Developer Experience

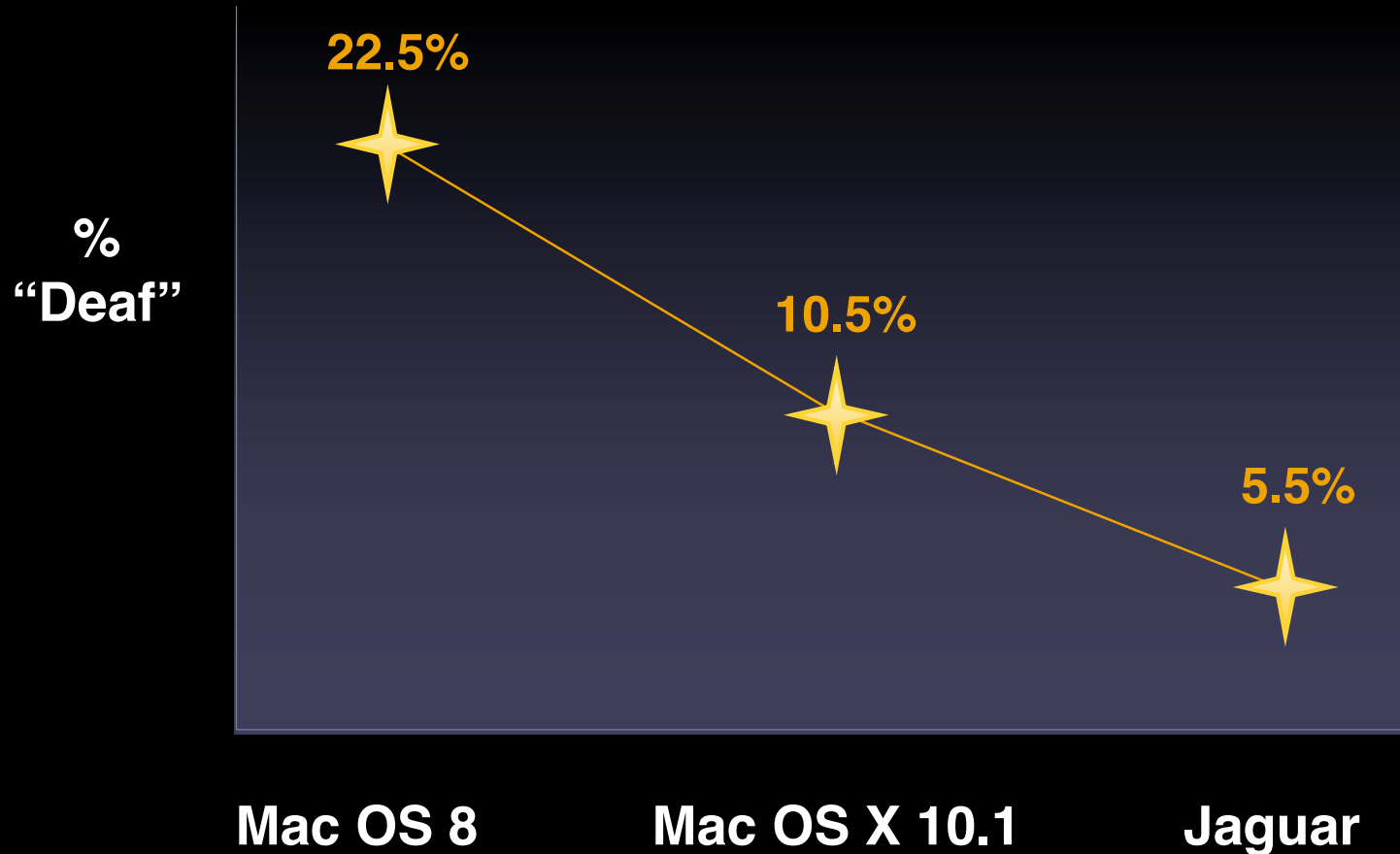


Improvements: Recognition

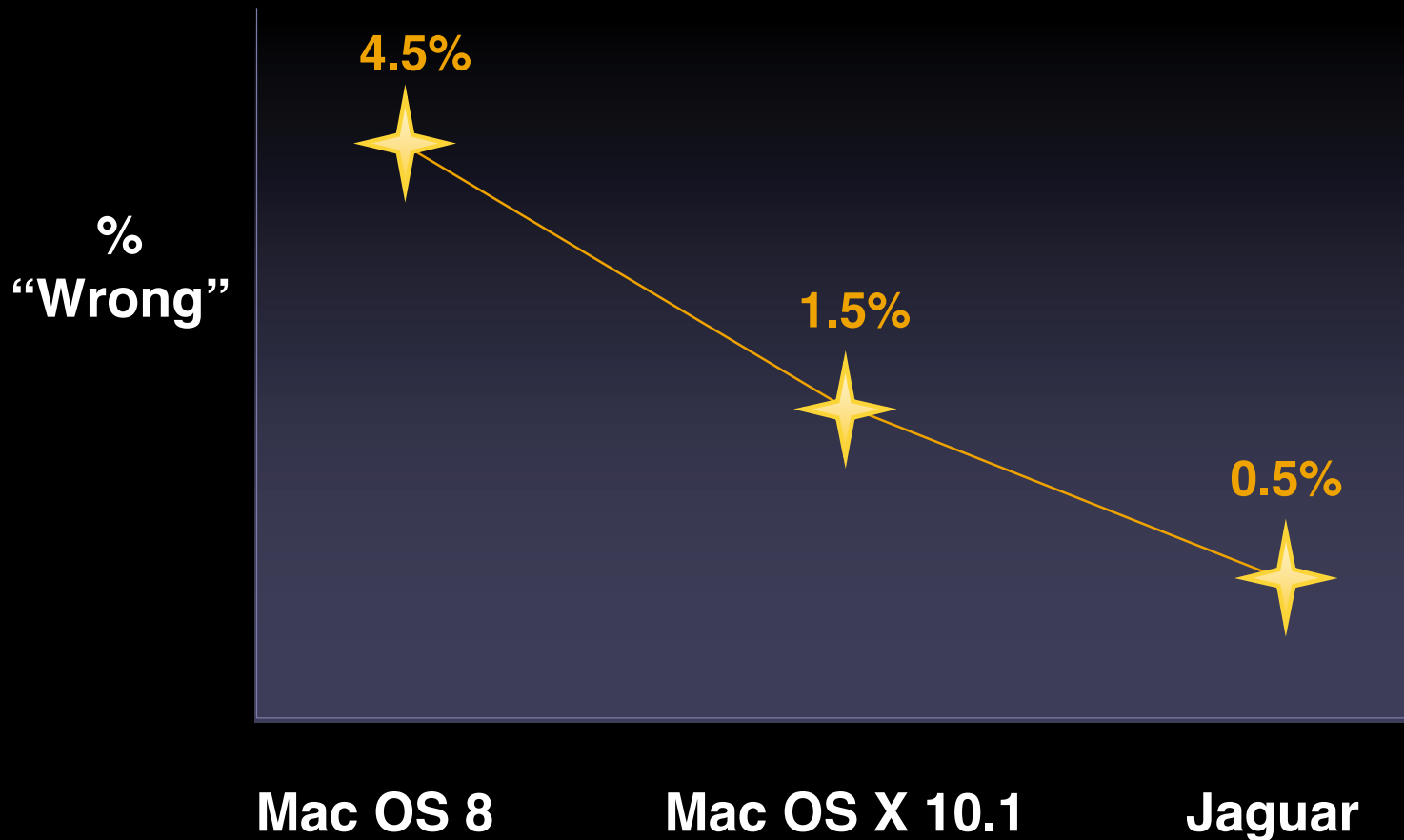
- Accuracy
- Robustness against background noise
- Continued optimization of microphones in iMacs, PowerBooks, and iBooks
- Improved compatibility with third party USB microphones



Speech Recognition Acceptance



Speech Recognition Accuracy



Improvements: Synthesis

- Better quality
 - More natural sounding
 - Easier to understand
- Improved voice signal processing
- New pitch, intonation, and rhythm models
- Fewer pronunciation errors
- Paragraph intonation improvements



Topic Tracking

“As computers have grown more powerful and their functions more various, the size of computer programs has grown dramatically. Because of their large size, many modern programs are distributed on CD-ROM’s.”

From World Book Encyclopedia, Mac OS X Edition



Discontinuities

*“Without these problems go
at those targets get the news.”*



Word Pronunciation

*“Brazilians and Ethiopians with Alzheimer’s
get their email with Outlook via PPPoE”*



Neologisms (New Words)

Myfile.cp

Readme.txt

Smileyface

Applelovers

Jobsification

Georgebushforpresident.info

Someunaccrediteduniversity.edu/phonenumbers.html



Improvements: User Experience

- Screen reader is built in (for free)
 - Reads any text on screen including any selected text
- Users can navigate the entire system (for free)
 - Menus and the front window including controls
- Talking notification
 - Background apps can do more than just jump in the Dock





Demo

Feature Improvements

Improvements: Developer

- TUNE Format
- Output to Audio File
- Embedded Speakable Commands
- Cocoa Classes



TUNE Format

- aka “Repeat after me”
- Specify tune and timing details
- Bypass the front end of the synthesizer
- Huge developer request



I like the good old nuts and bolts approach.

Pro Voice

New Modeling

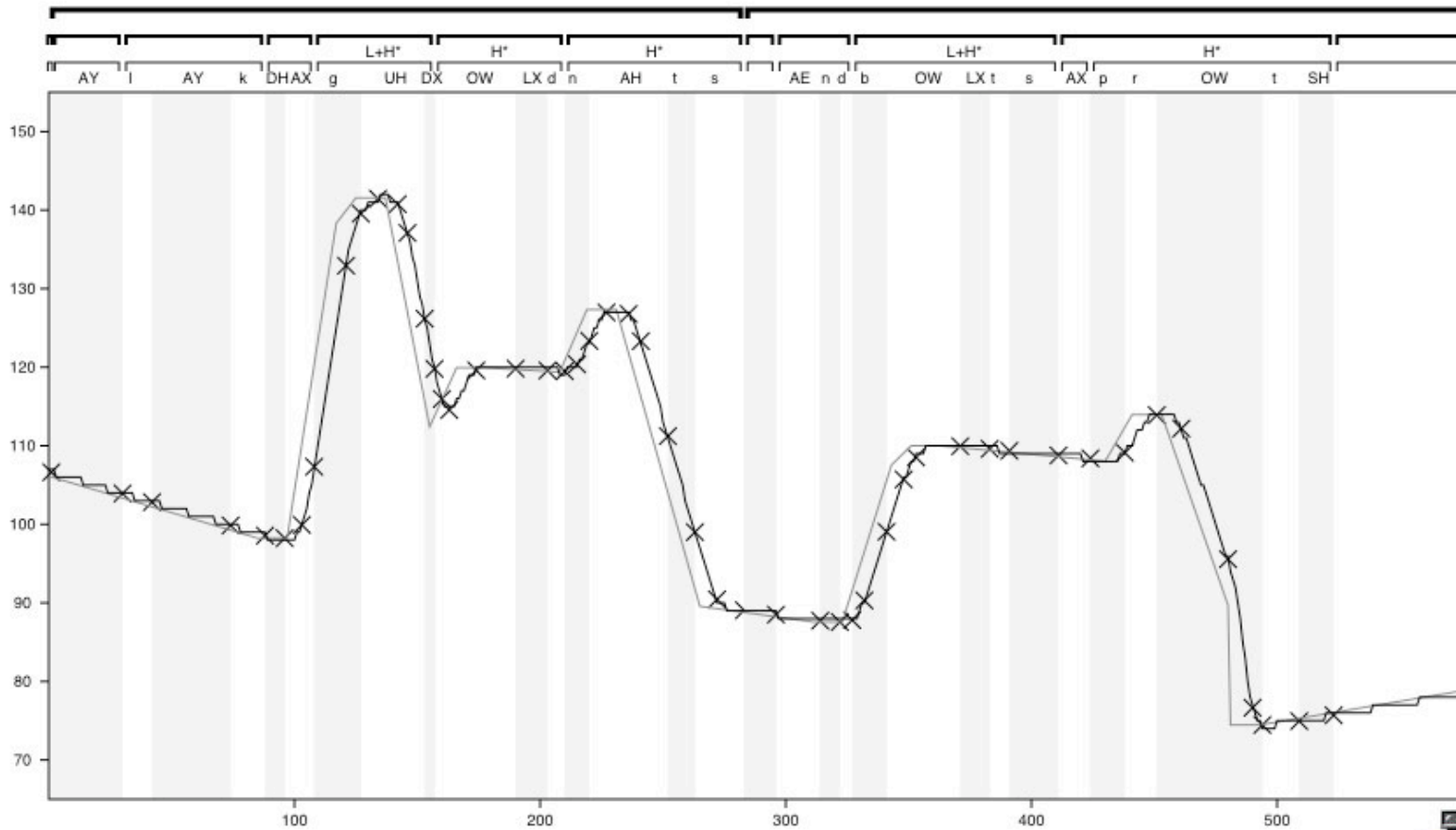
Duration

Pitch

Paragraph

Start

End



TUNE Format

AY <D 135; P 202.4:0>

l <D 55; P 197.3:0>

AY <D 150; P 195.2:0>

k <D 70; P 189.6:0>

D <D 40; P 187:0>

AX <D 55; P 186.7:0 187.6:27 191.7:55>

g <D 90; P 207.4:0 251.8:67 259.8:83>

UH <D 125; P 264.2:0 267.8:28 266.3:60 258.6:76>

d <D 20; P 239:0>



Output to Audio File

- Now released
- Example usage
 - Read email to audio file and download to iPod
- Developer opportunity to make new tools





Demo

TTS Output to Audio

Improvements: Developer

- TUNE Format
- Output to Audio File
- Embedded Speakable Commands
- Cocoa Classes



Cocoa Speech Classes

- Developer Request
- Greatly simplifies adoption and coding
- Prepares for the future



Cocoa Speech Classes

- Two main classes
 - Speech synthesis
 - Speech recognition



Cocoa Speech Synthesis

- One class
 - Can speak any string synchronously
 - Really easy!
- Need asynchronous or more control?
 - Setup a delegate for notification



Cocoa Speech Recognition

- Simple case uses one class
 - Get shared instance of class
 - Specify which commands you want to listen to
 - Begin listening



Cocoa Speech Recognition

- More sophisticated case
 - Handful of classes to build language model
 - Define what the user sees
 - Determine what notification you want
 - Start listening

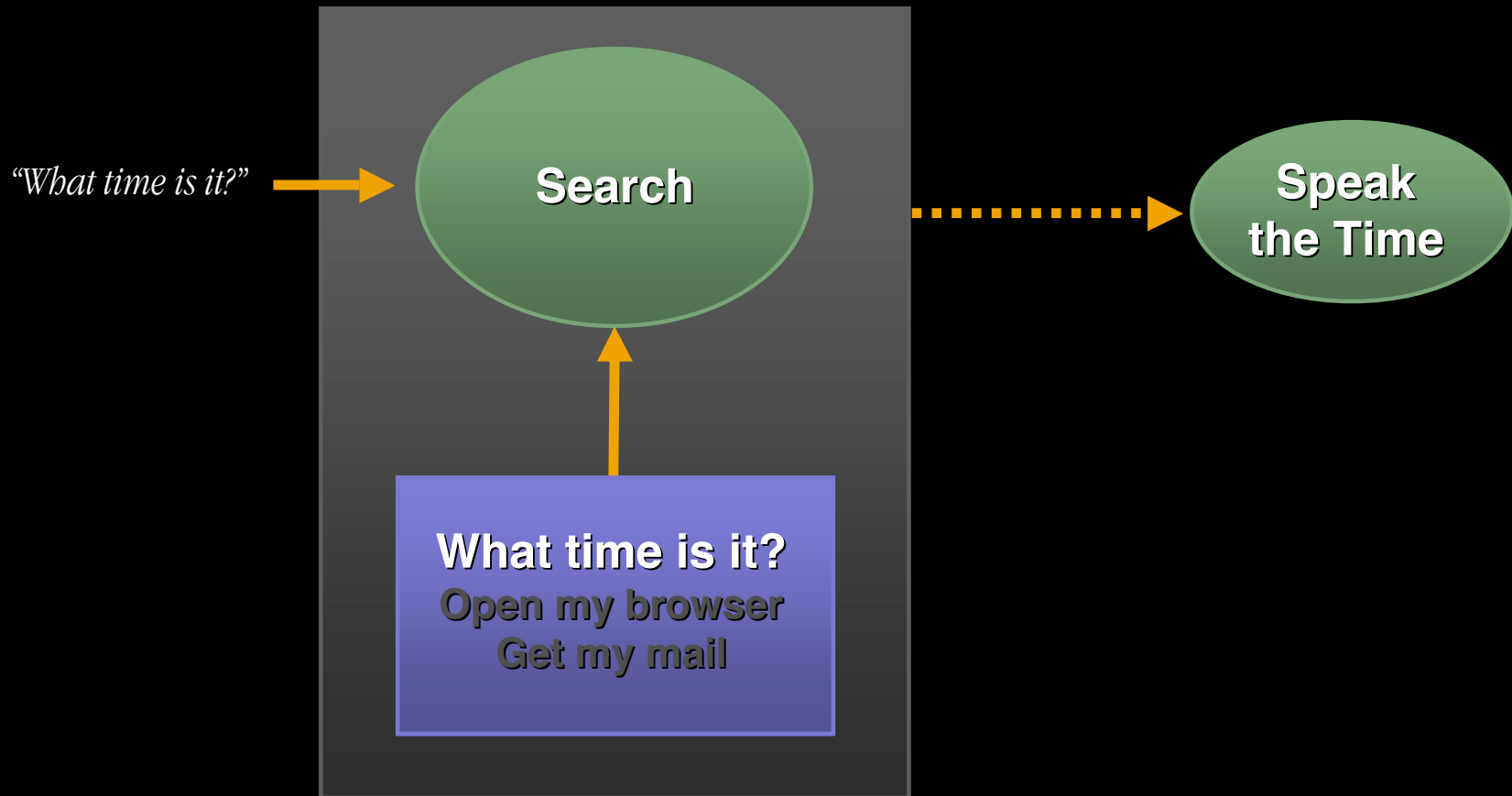


Improvements: User Experience

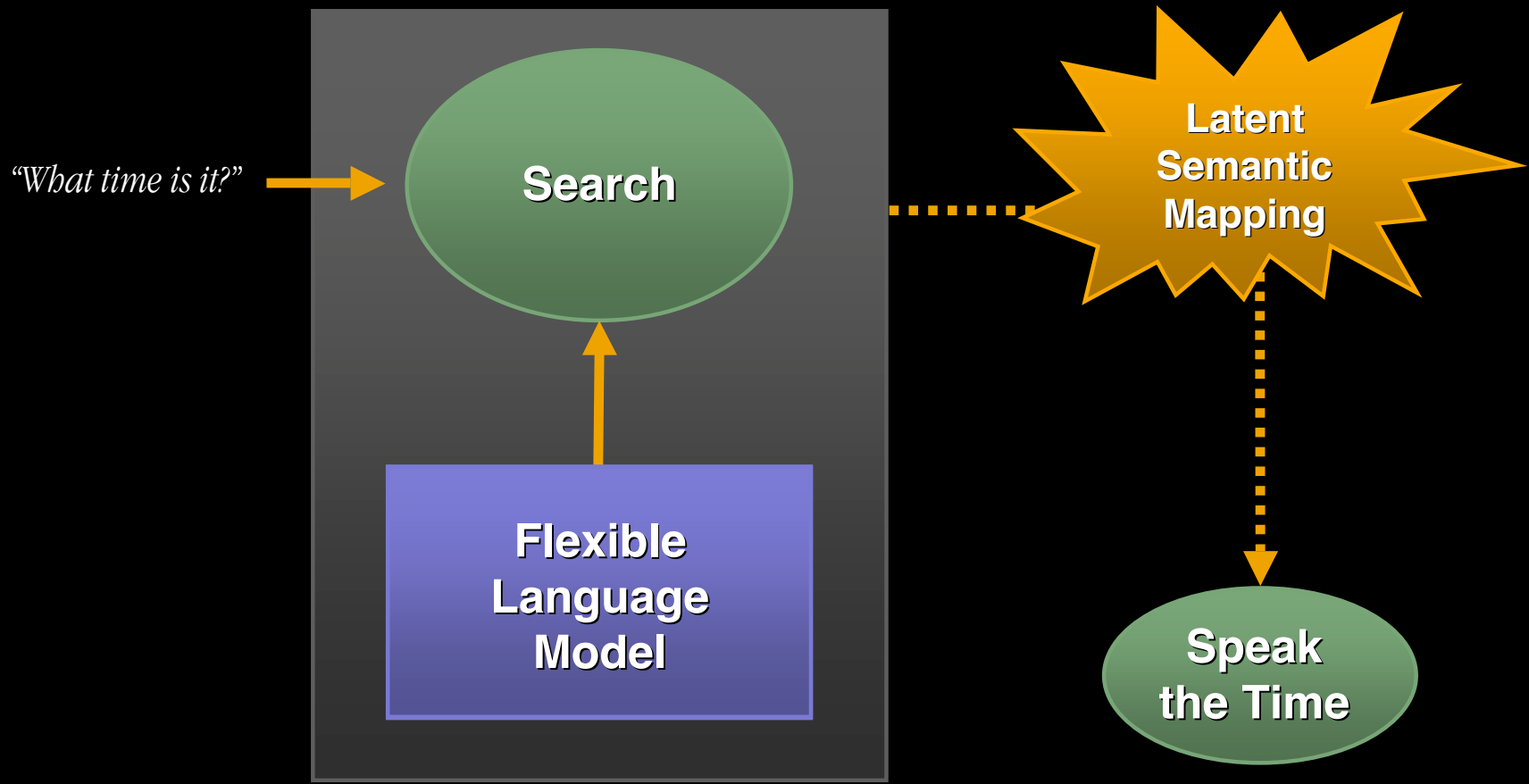
- Accessibility
- New Speakable Items
- Semantic Inference (“say what you mean”)



Semantic Inference



Semantic Inference





Demo

Semantic Inference

Tom Bonura

Deciding What To Do First.



Adopting Speech Recognition

Good

Better

Best



Adopting Speech Recognition

Use System Controls

Keyboard Shortcuts

Test with Speech
Turned On

Good

Better

Best



Test With Speech Turned On

- Turn on Speakable Items via the Speech System Preference panel
- Open the Speech Commands window and review contents
- Speak the name of every corresponding item in your UI



Keyboard Shortcuts

- Ensure menu items are associated with keyboard shortcuts
 - Focus on “buried” items
 - Focus on action “modifier” items
 - E.g., Hands-busy

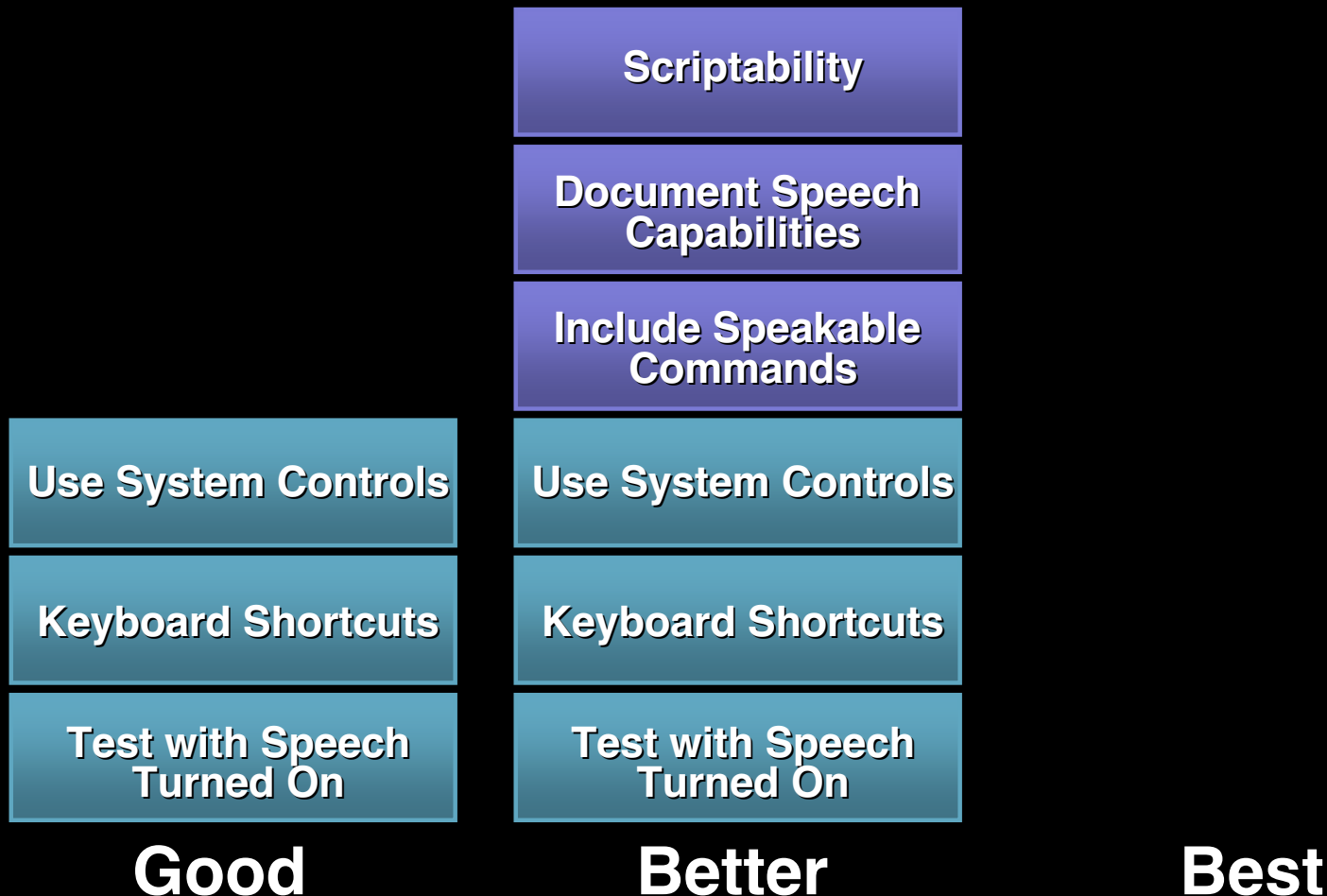


Use System Controls

- Avoid using custom controls
- If using custom controls
 - Follow accessibility guidelines



Adopting Speech Recognition



Incorporate Speakable Commands

- Allow Speakable Items to support you!
- Embed command files in your application (New)
 - (Any file that can be opened)
- Two file types are most helpful:
 - Compiled AppleScript files
 - XML-based command files
 - Use to send keyboard events to activate menus and controls via keyboard shortcuts



Document Speech Capabilities

- Bring to the users' attention that things are speakable
 - In your documentation
 - In your Help files
 - On your box
 - In the user interface
- Point users to Apple's Speech Tips in Mac Help



Scriptability

- Build upon AppleScript
 - Be scriptable
 - Be recordable
- Include speakable scripts
- Document your scriptability



Adopting Speech Recognition

Use Spoken Dialogs

Use Dynamic Language Models

Scriptability

Scriptability

Document Speech Capabilities

Document Speech Capabilities

Include Speakable Commands

Include Speakable Commands

Use System Controls

Use System Controls

Use System Controls

Keyboard Shortcuts

Keyboard Shortcuts

Keyboard Shortcuts

Test with Speech Turned On

Test with Speech Turned On

Test with Speech Turned On

Good

Better

Best



Use Dynamic Language Model

- Change what your application listens for based on current context
 - Keyboard focus
 - Current selection
- Make it flexible and forgiving



Spoken Dialogs

- Delegate a goal for the computer
- Ask the user questions in a multi-stage interaction
 - “Setup a meeting”
 - “Play some music”
 - “Send a new email”
 - “Do my taxes”



Adopting Speech Recognition

Use Spoken Dialogs

Use Dynamic Language Models

Scriptability

Scriptability

Document Speech Capabilities

Document Speech Capabilities

Include Speakable Commands

Include Speakable Commands

Use System Controls

Use System Controls

Use System Controls

Keyboard Shortcuts

Keyboard Shortcuts

Keyboard Shortcuts

Test with Speech Turned On

Test with Speech Turned On

Test with Speech Turned On

Good

Better

Best



Adopting Speech Synthesis

Good

Better

Best



Adopting Speech Synthesis

**Use System Alert
Dialogs/Sheets**

**Use Standard
Controls**

**Test with Speech
Turned ON**

Good

Better

Best



Test With TTS Turned ON

- Use Speech System Preference panel to turn on
 - “Speak Text Under Mouse”
 - “Use Hot key”
- Select then listen to all text in your user interface
 - Make sure it gets spoken
 - Rephrase or punctuate if it sounds wrong



Use Standard Controls

- Avoid using custom controls
- If using custom controls:
 - Support the “AppleSynthesis” accessibility attribute for custom controls
 - Provide separate speakable versions of titles etc., if necessary
 - Details to come



Use Standard Alert/Dialog Sheets

- Use only standard alert calls in Cocoa and Carbon



Adopting Speech Synthesis

Good	Better	Best
Use System Alert Dialogs/Sheets	Read Back Information	
Use Standard Controls	Customize Text	
Test with Speech Turned ON	Use System Alert Dialogs/Sheets	
	Use Standard Controls	
	Test with Speech Turned ON	



Customize Texts

- Customize wording
- Customize pronunciation
 - Use phonemes for mispronounced words
- Customize intonation
 - Let user catch up
 - De-emphasize familiar things
 - Liven it up!
 - Focus the listener's attention on the important elements
 - Use Paragraph Intonation



Read Back Information

- Spoken confirmation of what the computer is doing
 - Spreadsheet data
 - Calculator display
 - “Opening the pod bay doors”



Adopting Speech Synthesis

Use System Alert Dialogs/Sheets	Read Back Information	TUNE Input
Use Standard Controls	Customize Text	Spoken Notification
Test with Speech Turned ON	Use System Alert Dialogs/Sheets	Read Back Information
Good	Better	Best



Spoken Notification

- Notify users about asynchronous events
 - “Your Apple stock just hit \$65”
 - “It’s 3pm”
- Use Speech in lieu of `SystemBeep()`
 - “Your download has finished”
 - “Your printer is out of paper”
 - “You have a meeting with John in 5 minutes”



TUNE Format

- Mentioned earlier on



Adopting Speech Synthesis

Use System Alert Dialogs/Sheets	Read Back Information	TUNE Input
Use Standard Controls	Customize Text	Spoken Notification
Test with Speech Turned ON	Use System Alert Dialogs/Sheets	Read Back Information
Good	Better	Best



Issues

- Educating users how to speak
 - No pauses, don't shout, . . .
 - Refer to “Helpful Tips” in Speech preference panel
- Background noise
 - Head-mounted microphones
- Non-native speakers
- Localization



PlainTalk@apple.com



Documentation

Speech

- Speech Recognition Manager Reference

Documentation > Carbon > Multimedia > Speech Recognition Manager
developer.apple.com/techpubs/macosx/Carbon/multimedia/SpeechRecognitionManager/speechrecognition.html

- Speech Synthesis Manager Reference

Documentation > Carbon > Multimedia > Speech Synthesis Manager
developer.apple.com/techpubs/macosx/Carbon/multimedia/SpeechSynthesisManager/speechsynthesis.html



More Documentation

Speech

- Inside Macintosh: Speech Recognition Manager

[developer.apple.com/techpubs/mac/speechrecogmgr/
srec-2.html](http://developer.apple.com/techpubs/mac/speechrecogmgr/srec-2.html)

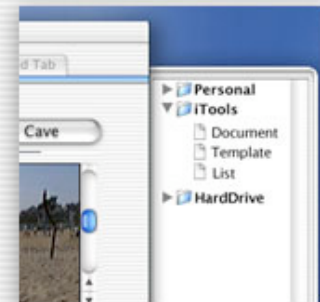
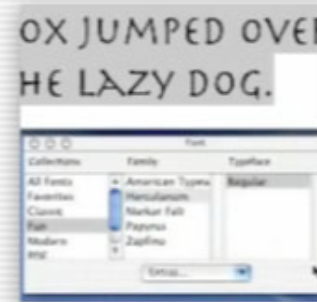
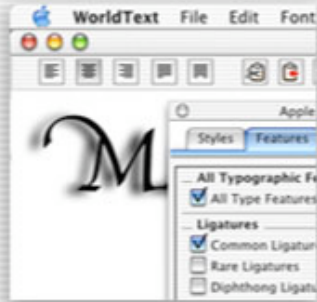
- Inside Macintosh: Sound
(Chapter 4, Speech Manager)

developer.apple.com/techpubs/mac/Sound/Sound-187.html





Q&A



John Geleynse
User Experience Evangelist
geleynse@apple.com

<http://developer.apple.com/wwdc2002/urls.html>

 **WWDC2002**

 **WWDC2002**

 **WWDC2002**